

## NEW TECHNIQUES FOR HIGH-CONTRAST IMAGING WITH ADI: THE ACORNS-ADI SEEDS DATA REDUCTION PIPELINE

TIMOTHY D. BRANDT<sup>1</sup>, MICHAEL W. McELWAIN<sup>2</sup>, EDWIN L. TURNER<sup>1,3</sup>, L. ABE<sup>4</sup>, W. BRANDNER<sup>5</sup>, J. CARSON<sup>6</sup>, S. EGNER<sup>7</sup>, M. FELDT<sup>8</sup>, T. GOLOTA<sup>7</sup>, M. GOTO<sup>3</sup>, C. A. GRADY<sup>2</sup>, O. GUYON<sup>7</sup>, J. HASHIMOTO<sup>9</sup>, Y. HAYANO<sup>7</sup>, M. HAYASHI<sup>9</sup>, S. HAYASHI<sup>7</sup>, T. HENNING<sup>5</sup>, K. W. HODAPP<sup>10</sup>, M. ISHII<sup>7</sup>, M. IYE<sup>9</sup>, M. JANSON<sup>1</sup>, R. KANDORI<sup>9</sup>, G. R. KNAPP<sup>1</sup>, T. KUDO<sup>9</sup>, N. KUSAKABE<sup>9</sup>, M. KUZUHARA<sup>9,11</sup>, J. KWON<sup>9,12</sup>, T. MATSUO<sup>9</sup>, S. MIYAMA<sup>13</sup>, J.-I. MORINO<sup>9</sup>, A. MORO-MARTÍN<sup>14</sup>, T. NISHIMURA<sup>7</sup>, T.-S. PYO<sup>7</sup>, E. SERABYN<sup>15</sup>, H. SUTO<sup>9</sup>, R. SUZUKI<sup>9</sup>, M. TAKAMI<sup>16</sup>, N. TAKATO<sup>7</sup>, H. TERADA<sup>7</sup>, C. THALMANN<sup>17</sup>, D. TOMONO<sup>7</sup>, M. WATANABE<sup>18</sup>, J. P. WISNIEWSKI<sup>19</sup>, T. YAMADA<sup>20</sup>, H. TAKAMI<sup>7</sup>, T. USUDA<sup>7</sup>, M. TAMURA<sup>9</sup>

*Draft version September 17, 2012*

### ABSTRACT

We describe Algorithms for Calibration, Optimized Registration, and Nulling the Star in Angular Differential Imaging (ACORNS-ADI), a new, parallelized software package to reduce high-contrast imaging data, and its application to data from the SEEDS survey. We implement several new algorithms, including a method to centroid saturated images, a trimmed mean for combining an image sequence that reduces noise by up to  $\sim 20\%$ , and a robust and computationally fast method to compute the sensitivity of a high-contrast observation everywhere on the field-of-view without introducing artificial sources. We also include a description of image processing steps to remove electronic artifacts specific to Hawaii2-RG detectors like the one used for SEEDS, and a detailed analysis of the Locally Optimized Combination of Images (LOCI) algorithm commonly used to reduce high-contrast imaging data. ACORNS-ADI is efficient and open-source, and includes several optional features which may improve performance on data from other instruments. ACORNS-ADI is freely available for download at [www.github.com/t-brandt/acorns-adi](http://www.github.com/t-brandt/acorns-adi) under a BSD license.

*Subject headings:*

### 1. INTRODUCTION

Since 1992, more than 700 confirmed exoplanets and 2000 additional candidates have been discovered<sup>22</sup>. Ground-based surveys have confirmed hundreds of exoplanets by measuring the periodic radial velocity shifts they induce in their host stars

(e.g., Vogt et al. 2000; Queloz et al. 2000; Tinney et al. 2001; Mayor et al. 2003) or by measuring photometric variations as they transit their host stars (e.g., Alonso et al. 2004; Bakos et al. 2004; McCullough et al. 2005; Pollacco et al. 2006; Charbonneau et al. 2009). In space, NASA's *Kepler* satellite (Borucki et al. 2010) has identified more than 2000 candidate transiting exoplanets. These indirect methods are sensitive to short-period exoplanets: the magnitude of a radial velocity signal and the probability of a transit both decrease with separation. These methods also generally require observations over several orbital periods, making them impractical for detecting exoplanets with periods of more than a few years.

Direct imaging surveys, made possible by advances in adaptive optics, infrared detectors, and image processing algorithms, are now complementing transit and radial velocity surveys, identifying giant exoplanets tens of astronomical units (AU) from their host stars. Ground-based high-contrast imaging surveys have shown that these giant exoplanets are rare (e.g., Biller et al. 2007; Lafrenière et al. 2007a), and are beginning to constrain models of exoplanet and exoplanetary system formation and evolution (Janson et al. 2012).

Large-scale direct-imaging surveys rely on sophisticated image processing to search for faint companions around bright stars. In addition to the usual bias, flat-field, and distortion corrections, these surveys must model and subtract the stellar point-spread function (PSF). Most surveys use Angular Differential Imaging (ADI) to make this task easier. As the Earth rotates, the orientation of the field-of-view (FOV) of an altitude-azimuth telescope (and thus of the PSF on the detector) changes relative to the celestial north. Features of the

\* Based on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan.

<sup>1</sup> Department of Astrophysical Sciences, Princeton University, Princeton, USA.

<sup>2</sup> Goddard Space Flight Center, Greenbelt, USA.

<sup>3</sup> Institute for the Physics and Mathematics of the Universe, University of Tokyo, Japan.

<sup>4</sup> Laboratoire Hippolyte Fizeau, Nice, France.

<sup>5</sup> Max Planck Institute for Astronomy, Heidelberg, Germany.

<sup>6</sup> College of Charleston, Charleston, South Carolina, USA.

<sup>7</sup> Subaru Telescope, Hilo, Hawai'i, USA.

<sup>8</sup> Universitäts-Sternwarte München, Munich, Germany

<sup>9</sup> National Astronomical Observatory of Japan, Tokyo, Japan

<sup>10</sup> Institute for Astronomy, University of Hawai'i, Hilo, Hawai'i, USA.

<sup>11</sup> University of Tokyo, Tokyo, Japan.

<sup>12</sup> Department of Astronomical Science, Graduate University for Advanced Studies, Tokyo, Japan

<sup>13</sup> Hiroshima University, Higashi-Hiroshima, Japan

<sup>14</sup> Department of Astrophysics, CAB - CSIC/INTA, Madrid, Spain.

<sup>15</sup> Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.

<sup>16</sup> Institute of Astronomy and Astrophysics, Academia Sinica, Taipei, Taiwan.

<sup>17</sup> Anton Pannekoek Astronomical Institute, University of Amsterdam, Amsterdam, The Netherlands.

<sup>18</sup> Department of CosmoSciences, Hokkaido University, Sapporo, Japan.

<sup>19</sup> HL Dodge Department of Physics and Astronomy, University of Oklahoma, Norman, OK, USA.

<sup>20</sup> Astronomical Institute, Tohoku University, Sendai, Japan

<sup>22</sup> See <http://www.exoplanet.eu/>

PSF due to the instrument, such as telescope spiders and the diffraction pattern, appear to rotate relative to any faint companion.

The first algorithms to take advantage of ADI used simple techniques to model the PSF, like taking the median of a sequence of exposures (Marois et al. 2006). More recently, algorithms like the Locally Optimized Combination of Images (LOCI, Lafrenière et al. 2007b) model the PSF locally, while principal component analysis (PCA)-based techniques (Soummer et al. 2012; Amara & Quanz 2012) model it globally. These more sophisticated algorithms can offer a factor of  $\sim 2$  or more improvement in sensitivity over simple ADI reductions.

In this paper, we present Algorithms for Calibration, Optimized Registration, and Nulling the Star in Angular Differential Imaging (ACORNS-ADI), a software package to analyze ADI data for the SEEDS survey (see Tamura 2009), a five-year direct imaging survey using the HiCIAO instrument (Hayano et al. 2008) on the Subaru Telescope. We discuss each non-trivial step of the reduction process, from the bias and flat-field corrections to PSF modeling to the final sensitivity analysis. ACORNS-ADI is parallelized, open-source, and freely available for download at [www.github.com/t-brandt/acorns-adi](http://www.github.com/t-brandt/acorns-adi) under a BSD license.

## 2. ADI DATA REDUCTION IN SEEDS

In order to take advantage of the field rotation in ADI, a series of short exposures is taken, with minimal field rotation during each individual exposure. The central star is usually allowed to saturate in order to increase the observing efficiency and limit the amount of read noise for a given number of companion photons. A typical high-contrast ADI dataset thus consists of a series of short, sequential exposures with a saturated central star. The data reduction process searches for point sources in the image sequence. While it is also possible to analyze extended structures, like disks, using ADI reduction techniques (e.g. Liu 2004; Thalmann et al. 2011), it is far more difficult to interpret the final processed images (see Section 5.1). A detailed discussion of extended sources in ADI is beyond the scope of this paper.

A reduction of a high-contrast ADI image sequence proceeds in several steps:

1. Correct for the bias, flat-field each image;
2. Correct the image distortion (if necessary);
3. Register the frames (if necessary);
4. Model and subtract the PSF of the central star;
5. Rotate each frame to align it to the celestial north;
6. Combine the sequence of images;
7. Search for point sources; and
8. Produce a sensitivity map.

Each step in the reduction impacts the sensitivity of the final combined image. Optimizing and characterizing this sensitivity is critical for understanding the incidence and properties of substellar companions.

Some of the steps listed above, such as the distortion correction (once the distortion map is known!) and the rotation to a common frame, are trivial. Others are surprisingly difficult, or can be optimized to give significant improvements in sensitivity. For example, a one-pixel root-mean-square scatter in the image registration can degrade sensitivity by  $\sim 20\%$  (Section 5.2), as can the use of the median intensity, rather than a trimmed mean, to combine a sequence of images (Section 6.1).

Figure 1 shows a sample SEEDS dataset through the above sequence of steps as processed by ACORNS-ADI. The first frame shows the central  $5'' \times 5''$  of a  $20'' \times 20''$  sample raw image, while the second frame shows the effect of correcting for the bias (Section 3.1), flat-fielding, and hot pixel masking (Section 3.2). The third frame shows the same image after correcting for field distortion (Section 3.3) and registering to the PSF centroid (Section 4). The first frame on the second line shows the residuals after subtracting the stellar PSF using the LOCI algorithm (Section 5.2, Appendix A); the next frame shows the combined image from an ADI sequence (Section 6.1, Appendix B) convolved with a circular  $0''.05$  aperture (Section 6.2) and normalized by the radial profile of its standard deviation. Finally, we show a radial profile of the dataset’s sensitivity to point sources (Section 6.3, Appendix A).

In the following sections, we describe each non-trivial step listed above. Some steps, like the subtraction of the stellar PSF and the combination of an image sequence, apply generally to data from any survey, while other steps, like the distortion correction and bias correction, have features specific to SEEDS data. Our discussions of two of these steps, the computation of a sensitivity map and the statistics of a combination of images, contain calculations that we relegate to appendices.

## 3. CALIBRATION

The first step in the data reduction is calibration: finding the count rate corresponding to zero intensity, flat-fielding, and applying a distortion correction. The zero point correction and flat-fielding are complex and inter-related for SEEDS data, and we handle them with a single routine. We then apply a distortion correction to the intensity-calibrated data. Because an ADI sequence consists of many short exposures, its sensitivity far from the central star can be limited by calibration uncertainties and read noise. Typical SEEDS observations are read noise limited at separations  $\gtrsim 2''$ .

The algorithms and discussions in this section are mostly specific to data from a  $2048 \times 2048$  pixel Hawaii2-RG (H2RG) detector (Blank et al. 2011). These HgCdTe detectors are becoming more common on major instruments and telescopes. In addition to HiCIAO on Subaru, Calar Alto, CFHT, VLT, IRTF, Keck, and SALT all use H2RG detectors. Future space-based missions such as *JWST*, *JDEM*, *EUCLID*, and Prime Focus Spectrograph and CHARIS, the next-generation spectrograph and camera for Subaru, will all contain H2RG or similar, but larger,  $4096 \times 4096$  pixel H4RG arrays. To reduce data from another detector, the user would need to replace this calibration module with one specific to his or her instrument.

### 3.1. Removing the Bias

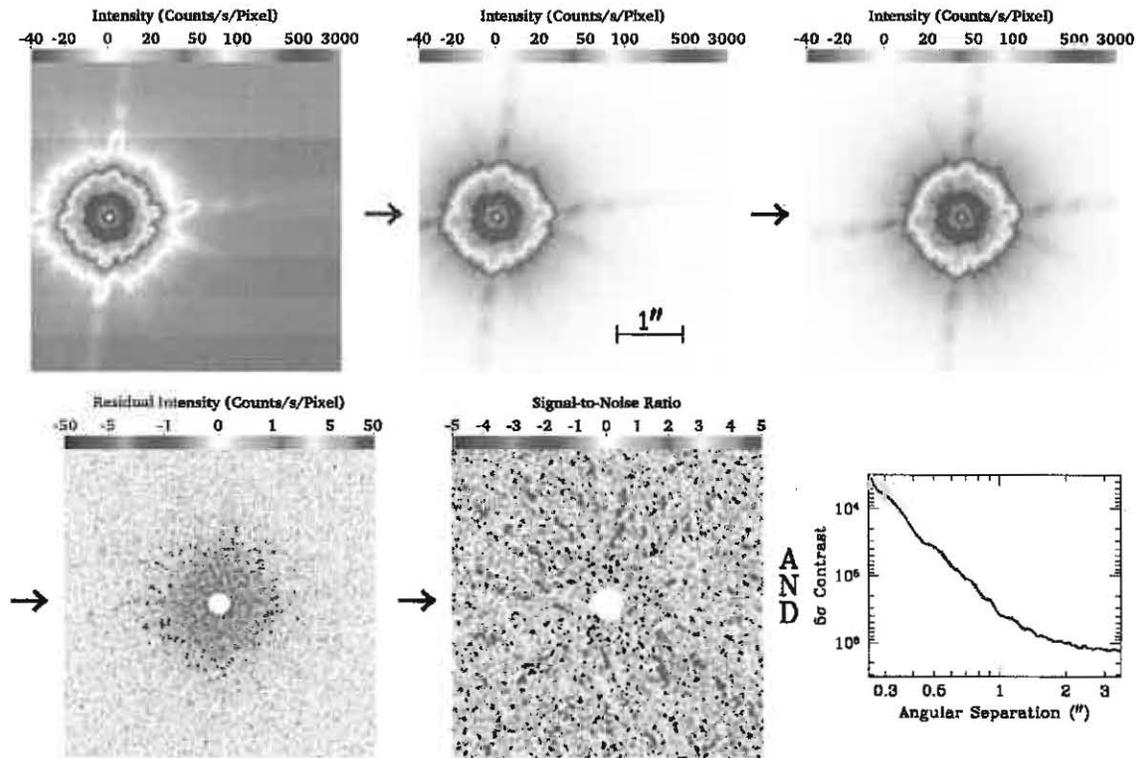


FIG. 1.— A step-by-step depiction of the ACORNS-ADI data reduction process, as discussed in Section 2. The first image shows the central  $5'' \times 5''$  of a  $20'' \times 20''$  frame of raw data. The second frame has been bias-corrected, flat-fielded, and had its hot pixels masked (Sections 3.1 and 3.2), while the third frame has been corrected for field distortion (Section 3.3) and registered to the PSF centroid (Section 4). The first frame on the second row shows the residuals in a single frame after applying the LOCI algorithm (Section 5.2, Appendix A), while the next frame shows the combined image from an ADI sequence (Section 6.1, Appendix B) convolved with a circular  $0''.05$  aperture (Section 6.2) and normalized by the standard deviation at each radius. The final frame shows a radial profile, with shading to represent azimuthal scatter, of the final sensitivity map (Section 6.3, Appendix A).

As configured for SEEDS, HiCIAO's H2RG detector reads out data in 32 channels at a pixel rate of 100 kHz. Reading out all  $2048 \times 2048$  pixels thus takes about 1.3 seconds. As discussed by Moseley et al. (2010), each of the 32 readout channels has its own reference voltage, which appears as a bias—a non-zero count level corresponding to zero intensity—in raw HiCIAO data. Superimposed on this is a time-varying reference voltage which is largely shared between the 32 channels. To calibrate the count level corresponding to zero intensity, we therefore need to fit for 32 stable reference voltages and at least one function. The H2RG detector provides four rows of pixels at each detector edge, for a total of 32,704 reference pixels (the 'R' in H2RG). These pixels are not light-sensitive, but are subject to the same reference voltages as the rest of the array, and can be used to estimate the pixel-by-pixel bias. For some observations, SEEDS has taken data with the guide window capability (the 'G' in H2RG), reading out only a subarray of the detector. In this mode, there is a single readout channel (and reference voltage) and there are no reference pixels.

For SEEDS data taken in the normal 32-channel readout mode, the 512 reference pixels at the ends of the channel are sufficient to determine the stable reference voltages. The H2RG read noise is very nearly Gaussian, so we use a sigma-reject technique to calculate their offset from zero. ACORNS-ADI iteratively rejects  $3\sigma$  outliers,

takes the mean of the remaining reference pixels, and subtracts this value from all of the pixels in each readout channel. This calibration is good to approximately the read noise over  $\sqrt{512}$  (the number of reference pixels), or about  $1 e^-$  per frame. The first row of Table 1 shows the read noise as measured in a series of 30 dark frames taken in December 2010 with no bias corrections. The difference between the root variance within a readout channel ( $27 e^-$ ) and over the entire array ( $52 e^-$ ) is almost completely removed by a mean subtraction using only the 512 reference pixels, as shown in the second row of Table 1.

The time-varying reference voltage, largely shared among the 32 readout channels, is more difficult to subtract. Moseley et al. (2010) describe two techniques. One of these relies on reference pixels interspersed throughout the detector array, which would be difficult to implement in an imaging survey like SEEDS. The other technique saves the reference voltage in place of one of the 32 channels of science pixels. Appropriate weighting of the Fourier components of this reference voltage then provides a good estimate of the bias to be subtracted from each readout channel.

The H2RG detector has  $1/f$  noise that extends from the frame rate,  $\sim 1$  Hz, to a knee at  $\sim 3$  kHz, where the noise becomes uncorrelated and the optimal weighting of the Fourier components falls to zero (Moseley et al.

2010). SEEDS does not currently save the reference voltage, which would require the sacrifice of 1/32 of the FOV; we have therefore estimated the possible improvement in read noise by applying a high-pass filter, removing all read noise with a frequency  $\lesssim 3$  kHz from each channel. Removing all low-frequency noise reduces the total read noise by about 10%, from 27  $e^-$  to 25  $e^-$  (fourth row of Table 1).

Without the reference voltage, ACORNS-ADI implements two techniques to estimate and subtract the reference voltage; the user must select which to use. The first technique uses the reference pixels at the edges of channels 1 and 32, which provide eight measurements of the reference voltage every 64 pixels. We first remove outliers with sigma-rejection, and then subtract the convolution of reference pixels with a Gaussian, choosing the width of the Gaussian to optimize the reference subtraction as follows.

While the H2RG detector has  $1/f$  noise that extends up to a knee at  $\sim 3$  kHz, sparse sampling (and read noise) limit our ability to measure the high frequency noise. As shown in Table 1, a perfect suppression of the noise up to  $\sim 3$  kHz would decrease the read noise by  $\sim 10\%$ , or the variance by just under 20%. However, a poor estimate of the reference voltage can increase the noise. We can then estimate our fractional suppression of the read noise as

$$\frac{0.2}{\ln 3000} \int_{1 \text{ Hz}}^{\nu_{\max}} \frac{d\nu}{\nu} - \frac{1}{N}, \quad (1)$$

where  $N$  is the effective number of pixels used to estimate the reference voltage at each point,  $\nu_{\max} \propto N^{-1}$ , and  $\ln 3000$  is the value of the integral with  $\nu_{\max} = 3$  kHz. Maximizing Eq. (1), we find that  $N \sim 40$ . Since only one in eight pixels has a corresponding reference pixel, this is equivalent to an effective smoothing length of  $\sim 300$  science pixels. We therefore cannot suppress noise of  $\gtrsim 300$  Hz, and achieve a  $\sim 5\%$  reduction in read noise (third row of Table 1) rather than a  $\sim 10\%$  reduction (fourth row). Like Moseley et al. (2010), we optimally suppress low-frequency read noise by scaling the reference signal; we find a best-fit scaling of 0.87.

ACORNS-ADI also implements a technique to remove correlated read noise using some or all of the science pixels. To give an accurate estimate of the bias, these pixels must be uniformly illuminated. SEEDS generally observes a saturated central source whose seeing halo extends out to several hundred pixels in radius (several arcseconds); pixels beyond this halo can be used to better estimate the high frequency components of the reference voltage. ACORNS-ADI estimates the (uniform) illumination using the difference between the reference and science pixels, and takes the median of each set of up to 32 simultaneous readouts. Using all of the science pixels, this procedure reduces the read noise by  $\sim 10\%$  over the perfect subtraction of all read noise of frequency  $\lesssim 3$  kHz (last line of Table 1), much more than the  $\sim 2\%$  expected from self-subtraction. The result is nearly as good when using only half the science pixels, those at least 800 pixels from the center (fifth line of Table 1). This level,  $\sim 23 e^-$ , indicates the read noise that is not shared between readout channels.

### 3.2. Flat-Fielding and Hot Pixel Masking

TABLE 1  
AVERAGE RESIDUAL READ NOISE AFTER BIAS CORRECTION IN  
30 DARK FRAMES

Zero Point Method	Average Read Noise ( $e^-/\text{pixel}$ )	
	Single Channel	Entire FOV
None	27.2	51.8
One Voltage per Channel	27.2	27.4
Reference Pixels, $\nu < 300$ Hz	26.1	26.2
All Pixels, $\nu < 3$ kHz	25.0	25.1
Half of the Pixels, Median	23.4	23.8
All Pixels, Median	22.3	22.5

For a typical exposure of  $\sim 1$  to 10 seconds, the dark current is  $\sim 0.05$  to  $0.5 e^- s^{-1}$ , a factor of  $\sim 50$  to 500 lower than the read noise. Pixel-to-pixel variations in the dark current will be still lower. A perfect suppression of the dark current would reduce the read noise by  $\lesssim 0.01\%$ ; we therefore do not attempt to correct for it, but treat the dark current as part of the background.

HiCIAO shares the problem of “hot,” indium-contaminated pixels with other H2RG detectors. As a result,  $\sim 2\%$  of HiCIAO’s pixels are unusable. These hot pixels are stable from night-to-night, though because the detector slowly degrades with time, new hot pixel maps must be made periodically. ACORNS-ADI corrects these pixels by taking the median of all uncontaminated pixels in a  $5 \times 5$  box centered on the hot pixel. Because the data are oversampled, this will not significantly bias the intensity. We mask hot pixels throughout the bias calculation described above.

Because of the short exposure times in most SEEDS images, cosmic rays are rarely a problem. We do implement an algorithm to reject isolated discrepant pixels or clusters of pixels similar to the one used by Lafrenière et al. (2007b). We aggressively smooth the image with a large median filter, identify pixels that are above a certain threshold in the difference between the original and the smoothed maps, and mask them.

Flat-field images are stable to  $\sim 2\%$  from month to month, and are even more stable from night to night. We therefore construct one flat-field for each observing run from nine dithered dome flats. We correct the bias of each frame with the method described in Section 3.1 using only the reference pixels, and then median-combine the nine dithered images. ACORNS-ADI divides each science frame by this master flat after performing the bias correction.

### 3.3. Distortion Correction

The field distortion is the difference between positions on the detector and positions on-sky. We model the distortion using a third-order two-dimensional (2D) polynomial in pixel coordinates relative to the center of the FOV; the distortion at the image center is thus zero by definition. We use *Hubble Space Telescope* images of the globular clusters M5 and M15 for our reference images, extracting stellar positions using SExtractor (Bertin & Arnouts 2010). We use the same tool to extract stellar positions in HiCIAO images. We then fit for the polynomial coefficients, using Markov Chain Monte Carlo (Press et al. 2007) to minimize the difference between *Hubble* positions and corrected HiCIAO positions. This technique gives best-fit values and errors on all pa-

rameters. On a good night, we measure the horizontal and vertical pixel scales to 0.01%, and the direction of true north to a precision of  $\sim 0.005^\circ$ ; poor conditions degrade our precision by a factor of up to  $\sim 5$ . Finally, we use bilinear interpolation to transform our image to the new coordinate system, which has a pixel scale of 9.5 mas. The distortion correction is described in more detail by (?).

Our corrected images appear to be free of systematics, with the distribution of offsets between *Hubble* and HiCIAO positions being Gaussian in both the horizontal and vertical coordinates, and random over the FOV. The orientation of true north has been stable to  $\sim 0.03^\circ$  since the SEEDS survey began in 2009, but the plate scale has changed by up to 2% due to small changes in the optical setup and a new camera lens which was installed in April of 2011. Apart from changes in the overall plate scale, the field distortion has been extremely stable over the duration of the SEEDS survey. It is dominated by a  $\sim 3\%$  difference between the horizontal and vertical pixel scales and a  $\sim 0.3^\circ$  offset between the vertical axis on the detector and the celestial north.

#### 4. IMAGE REGISTRATION

Image registration is important both to optimize the subtraction of the stellar PSF and to maximize companion flux in the final de-rotated, co-added image. Unfortunately, it is difficult to centroid saturated stars, and most SEEDS data have no bright but unsaturated stars in the  $20'' \times 20''$  FOV. Here, we present a new algorithm to centroid isolated, saturated HiCIAO PSFs. This algorithm performs well, with a residual scatter of  $\sim 1$ – $2$  mas (0.1–0.2 pixels) for observations made in good conditions, and could easily be applied to data from other instruments.

##### 4.1. Developing a PSF Template

To accurately centroid images, we need to build a model of the HiCIAO PSF. Unfortunately, HiCIAO's AO system must be re-tuned at the beginning of each observation, and its PSF varies accordingly. Figure 2 shows this variation in a representative sample of SEEDS images. The top-left PSF is unsaturated, while the other panels each show a typical PSF of a different ADI sequence. The PSF varies strongly with atmospheric conditions and the performance of AO188, HiCIAO's 188-actuator adaptive optics system (Hayano et al. 2008), which itself depends on stellar brightness. To capture this variation, we proceed empirically, ultimately building a set of three PSF templates from nearly 3000 individual exposures of saturated stars observed as part of the SEEDS survey.

We extract three templates from 3000 images by iteratively registering the frames and performing PCA. We initially centroid the frames by fitting Moffat profiles. We manually remove outliers—some of which result from bad data, others from the failure of the algorithm—and use a scaled, unsaturated PSF to flag saturated pixels. We then rescale the data to a common flux and perform PCA on the sequence of images. We use weighted PCA, estimating the noise at each unsaturated pixel as the sum of read noise and photon noise, and ignore saturated pixels.

We now use the mean PSF and the first two principal components from this first pass to refine our cen-

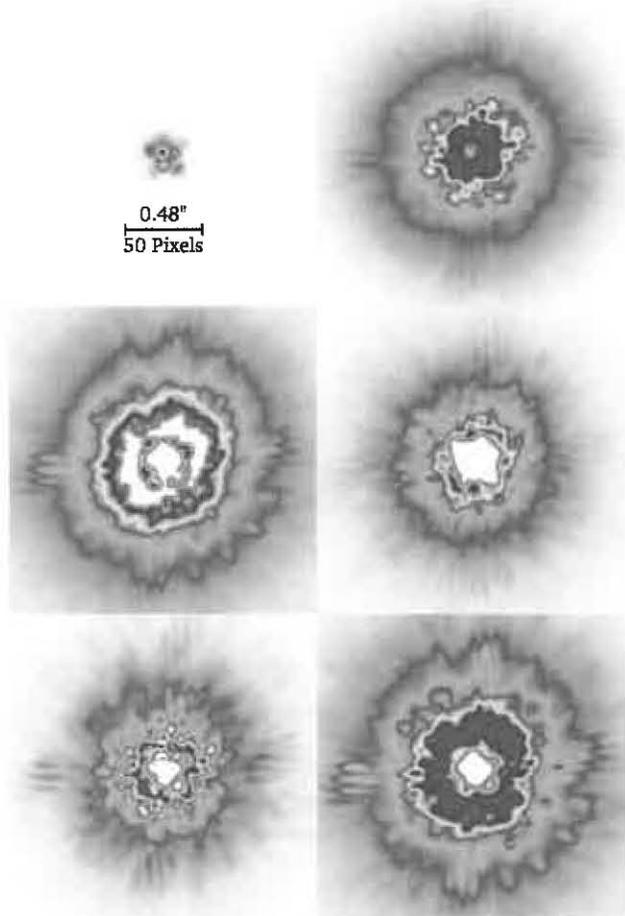


FIG. 2. — Sample HiCIAO PSFs. The top-left panel is unsaturated and includes the scale, while the other panels each represent a typical PSF of a different object observed as part of the SEEDS survey. The color scale is linear with white representing zero; because HiCIAO reads each pixel twice, areas that saturate rapidly can appear white. Using the algorithm described in Section 4, we can centroid each PSF to an accuracy of better than 0.5 pixels. For the better-behaved PSFs, like that shown the bottom left, our accuracy improves to  $\sim 0.1$ – $0.2$  pixels, or  $\sim 1$ – $2$  mas.

troids. We first re-estimate the centroid of each image by flagging saturated pixels (those with at least 70% of the maximum intensity on the detector) and centroiding the greatest concentration of such pixels. We then compute the root-mean-square distance  $r_{\text{rms}}$  of the saturated pixels from the provisional center, and mask all pixels within  $1.5r_{\text{rms}}$ . We model the variance of the intensity at each remaining pixel as shot noise plus read noise, and fit the frame with a linear combination of the mean PSF and the first two principal components by minimizing

$$\chi^2 = \sum_{\text{pix } i} \frac{1}{\sigma_i^2} \left( I_i - \sum_{\text{tmpl } j} \alpha_j T_{ij} \right)^2, \quad (2)$$

where  $j$  indexes the templates  $T$  ( $T_0$  represents the mean PSF),  $i$  indexes the pixels, and  $\alpha_j$  are free parameters. We then shift the template PSF by an integer number of pixels relative to the image, computing the best-fit  $\chi^2$  as a function of positional offset. Finally, we centroid the  $\chi^2$

map by fitting a two-dimensional quadratic, and take the location of this peak to be the centroid of the image. We then register all of the frames, scale them to a common flux, mask saturated pixels, weight the other pixels, and perform PCA. We repeat this entire sequence of steps one more time to build our final set of PSF templates, which we take to be the mean PSF and the first two principal components.

#### 4.2. Centroiding Saturated Frames

We centroid each sequence of saturated frames using the same method described in the previous section. The algorithm initially proceeds in four steps:

1. Flag saturated pixels, centroid the greatest concentration of such pixels to compute a provisional center;
2. Mask pixels near the provisional center, weight all other pixels by photon and read noise;
3. Fit the PSF templates using  $\chi^2$  at a grid of offsets; and
4. Centroid the map of  $\chi^2$  merit statistics.

We then recenter each frame and compute the average PSF of the ADI sequence. As Figure 2 shows, the average PSF in a given sequence of images can be significantly asymmetric. We therefore compute relative centroids according to the procedure just described, and separately determine the centroid of the average PSF.

A peculiar feature of HiCIAO makes it possible to visually estimate the centroid of a saturated frame. The HgCdTe H2RG chip is reset pixel-by-pixel and then read out twice; the interface computer records the difference between the two readings. As a result, pixels that saturate rapidly tend to show no difference between the two readings, and hence, are recorded as having zero intensity. We therefore allow the user to interactively select the absolute centroid of the average of an ADI sequence. Figure 3 shows a sample centroid verification image; the innermost circle has a radius of 4 HiCIAO pixels, or about 38 mas. As Figure 3 indicates, it is usually possible to absolutely centroid a sequence of saturated images to an accuracy of at least  $\sim 0.5$  pixels, around 5 mas. We then add this user-determined offset to each frame's centroid. In this way, ACORNS-ADI determines the relative centroids automatically, while the user selects the absolute centroid, in the form of a single offset for the sequence of images.

#### 4.3. Performance of the Centroiding Algorithm

We measure the performance of our new centroiding algorithm using ADI image sequences of single stars. The performance consists of the centroids' relative accuracy, which is completely determined by ACORNS-ADI, and their absolute accuracy, which is determined by the user and is much more difficult to assess. The accuracy of the relative centroids affects the quality of the data reduction, but not its astrometry. Poor image registration will smear out speckles and point sources but without introducing systematic offsets. The astrometric error in the reduced data is thus equal to the error in the absolute centroid, and must be estimated by the user. This varies

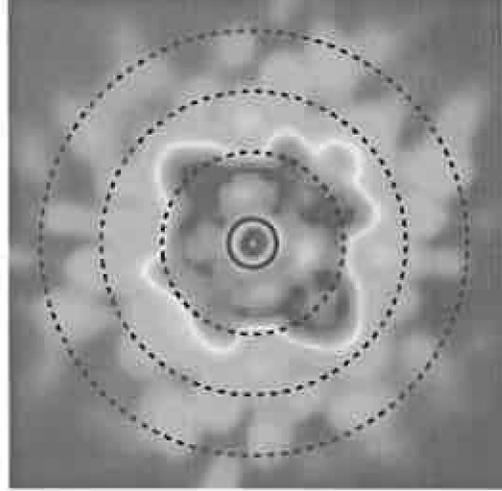


FIG. 3.— Sample average HiCIAO PSF from an ADI sequence. The user interactively chooses the absolute centroid in the average frame; the resulting offset is applied to each frame. The yellow cross shows the center of the image while the innermost circle has a radius of 4 HiCIAO pixels, or about 38 mas; the image is interpolated between points and therefore appears smooth. The absolute centroids in a typical sequence may be estimated to an accuracy of  $\lesssim 0.5$  pixels  $\approx 5$  mas; the relative centroids, as discussed in Section 4.3, are better.

from dataset to dataset but is generally  $\lesssim 0.5$  pixels,  $\sim 5$  mas.

We can assign upper limits to the errors in image registration using the scatter in fitted centroid positions. This scatter will be due both to tracking errors and PSF fitting errors, which we assume to be uncorrelated. We further assume that tracking errors dominate the slow drift in the PSF position. Before the installation of an atmospheric dispersion corrector (ADC, Egner et al. 2010) in AO188, the PSF would drift over the course of an observation due to differential atmospheric refraction between the visible, in which the AO system guides, and the near-infrared in which HiCIAO observes. This effect was mostly, though not entirely, eliminated by the installation of the ADC. More recent observations at high airmass (Thalmann et al. 2011) indicate that these slow drifts occur at the level of at most a few pixels (tens of mas) over the course of an ADI sequence.

As we wish to measure only the frame-to-frame fluctuations in the fitted centroids, we fit and subtract a low-order polynomial from the centroid positions in each ADI sequence. An alternative approach, measuring the positional difference between successive frames and dividing by  $\sqrt{2}$ , gives nearly identical results. We then compute the rms scatter of the residual centroid positions, excluding the most discrepant 1% of points to remove outliers. For a true Gaussian distribution, this outlier exclusion reduces the variance by 10%; we correct our measured variances for this effect. The worst datasets, those in which the PSF varies most and positional jitter is most likely to be real, show a root-mean-square (rms) scatter of  $\sim 0.3$  pixels (3 mas) in both the horizontal and vertical directions, or  $\sim 0.4$  pixels overall. Most of the data are much better; 12 of the 21 ADI sequences we used

to build the template PSFs showed overall residual rms scatters of less than 0.2 pixels (2 mas), and 17 of 21 had rms scatters of less than 0.3 pixels (3 mas).

Given the variation and asymmetries in HiCIAO's PSF, we allow the user to interactively determine the absolute centroid of an image sequence. However, the algorithm presented above can register a series of images to a typical precision of  $\sim 0.2$  pixels, or 2 mas. By centroiding a map of  $\chi^2$ , which itself is computed only at integer pixel offsets, our new method avoids interpolating images or PSF templates. This makes it relatively fast and free of systematics.

## 5. ADI REDUCTION

The goal of an ADI reduction process is to subtract the stellar PSF in a way that maximizes sensitivity to point sources in the residual images. In practice, this means finding an algorithm which produces Gaussian residuals and an optimal signal-to-noise ratio for single point sources.

We implement two basic techniques to model the PSF for each frame; these may be used alone or in conjunction with one another. The first method is to take the median of all of the frames to be the model PSF, subtract this from each individual image, and finally de-rotate and coadd the sequence. The second technique is Locally Optimized Combination of Images (LOCI), described by Lafrenière et al. (2007b). We discuss several modifications of the basic LOCI algorithm, some of which have been described elsewhere. Unfortunately, the AO system on the Subaru Telescope does not perform well enough in the  $H$ -band to take advantage of most of these techniques.

We characterize each data reduction algorithm by its *effective PSF*, which we define to be the difference between a reduced image with and without a faint point source. The effective PSF varies with the choice of data reduction algorithm and with position on the detector; it is a product both of hardware (the PSF itself) and of software. For most SEEDS datasets, we find that the basic LOCI algorithm offers the best compromise between sensitivity and simplicity. When SCExAO (Guyon et al. 2011), the new higher-order adaptive optics system for the Subaru Telescope, is fully operational, other algorithms may offer significant sensitivity improvements.

### 5.1. Median PSF Subtraction

A simple way to model the PSF is to use the median of all frames in an ADI sequence (c.f., Marois et al. 2006). The model PSF will then include all structures and companions in the FOV averaged over all position angles in the image sequence. Azimuthally extended sources will appear in the model PSF much as they do in individual exposures, while point sources will be smeared out by field rotation.

In this simple technique, the same model PSF is subtracted from each frame. The frames are then de-rotated to a common reference position and co-added. Variations in the PSF, such as a changing Strehl ratio, will strongly degrade the sensitivity of the final, processed image to point sources. A point source itself will suffer from a fractional flux loss roughly proportional to the ratio of the size of the PSF core to the amount of field rotation

at its location,

$$f_{\text{loss}} \sim \frac{\lambda}{D} \frac{1}{r_{\text{sep}} \Delta\phi} \quad (3)$$

where  $\Delta\phi$  is the total field rotation and  $r_{\text{sep}}$  is the angular separation between the point source and the central star. For a companion at  $r_{\text{sep}} \sim 1''$  with an  $H$ -band PSF core  $\lambda/D \sim 0''.05$  and a total field rotation of  $60^\circ$ ,  $f_{\text{loss}} \sim 5\%$ .

Azimuthally extended sources, like disks, will be suppressed by a factor

$$f_{\text{loss}}(\phi) \sim \int_{\phi-\Delta\phi/2}^{\phi+\Delta\phi/2} \frac{I(\theta) d\theta}{I(\phi)} \quad (4)$$

If we expand  $I$  as a Taylor series about  $\phi$ , all of the odd terms vanish due to the symmetry of the integral. In other words, azimuthal gradients, as well as azimuthally symmetric sources, are completely suppressed; only higher-order features survive a median PSF subtraction.

The left two panels of Figure 4 show the original PSFs in annuli, smeared out by the field rotation of the dataset, and the effective PSFs after a mean PSF subtraction. The latter are, on average, identical to the effective PSFs produced by a median PSF subtraction. The dataset shown had 155 exposures, with a total field rotation of  $\sim 30^\circ$ , and is typical of a SEEDS observation. The integrated flux in the mean PSF subtracted image is zero to within 0.1% of the flux in the raw PSF image.

Median PSF subtraction is simple both conceptually and computationally, and has been successfully used to measure the geometry of circumstellar disks (Thalmann et al. 2011). However, as discussed above, this technique preserves only high-azimuthal-order disk features; as a result, it is only effective on a small subset of the SEEDS disk sample. Furthermore, other techniques such as LOCI, described in the following section, are much more sensitive to point sources. We include median subtraction as an option in ACORN-ADI but generally recommend against its use. We use it here mainly as a baseline against which to measure the performance of other algorithms.

### 5.2. LOCI

LOCI, a technique for empirical PSF modeling, was introduced by Lafrenière et al. (2007b). The LOCI algorithm models the PSF in an ADI frame as a local linear combination of other frames in the sequence, with the coefficients calculated using simple least-squares. In each region of frame  $i$ , LOCI takes the other frames  $\{j\}$  and fits coefficients  $\{\alpha_{ij}\}$ , eventually producing an image of residual intensity,

$$\mathcal{R}_i = I_i - \sum_j \alpha_{ij} I_j \quad (5)$$

LOCI fits for the  $\{\alpha_{ij}\}$  over optimization regions typically several hundred PSF footprints—several thousand pixels—in size. We give more details about this calculation in Appendix A. Because LOCI uses least-squares fitting, the solution for the  $\{\alpha_{ij}\}$  is a linear problem, with the size of the resulting linear system set by the number of frames in an ADI sequence.

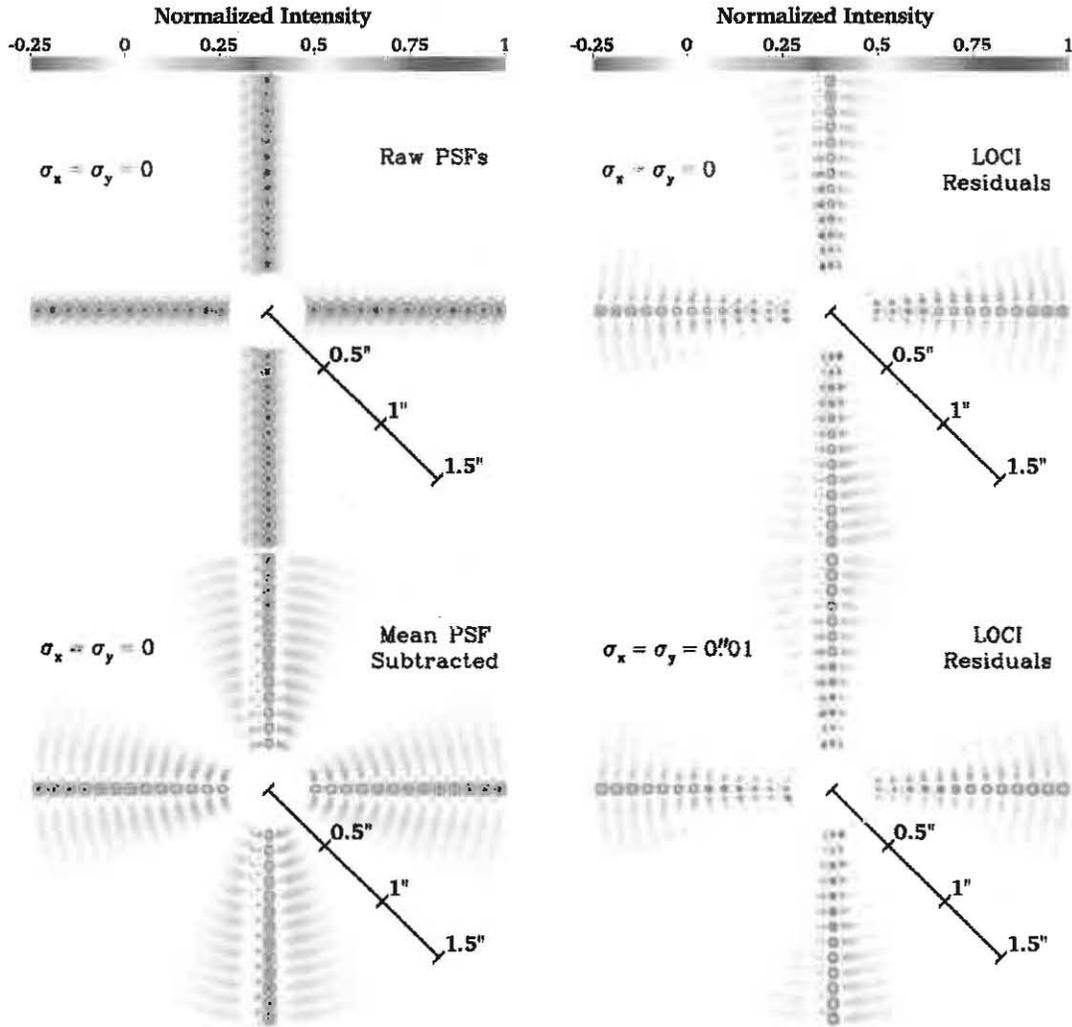


FIG. 4. — The effect of ADI image processing on the original H-band PSF (top left panel); all point sources in all panels were reduced independently of one another. We normalize the intensity to the peak in the original PSF. Lower-left panel: the effect of subtracting a mean companion PSF; for a faint source, this is equivalent to a median PSF subtraction (Equation (3)). The right panels show the residuals after running a LOCI reduction (Lafrenière et al. 2007b), with (bottom-right) and without (top-left) errors in imaging registration  $\sigma_x$  and  $\sigma_y$  prior to the reduction. A random Gaussian positional error of  $0''.01$ ,  $1/6$  of the PSF FWHM, reduces the flux in the effective PSF core by  $\sim 20\%$ . The integrated flux in each ADI-processed PSF is approximately zero. The LOCI-processed PSFs are difficult to characterize at small radii, and at all radii, they depend strongly on the precision and stability of the image registration.

To interpret a LOCI-processed ADI sequence, artificial point sources are added and the data are re-reduced (Lafrenière et al. 2007b). Here, we define the LOCI effective PSF to be the difference between the final, reduced image with and without a faint point source. The right panels of Figure 4 show the effective PSF after reducing a sample SEEDS dataset with LOCI. The dataset, with 155 frames and a total field rotation of  $\sim 30^\circ$ , represents a typical SEEDS observation. To ensure that each effective PSF is independent from the others, we add and reduce the faint companions one at a time. We use optimization regions 200 PSF footprints in size and a minimum field rotation of 70% of the PSF full width at half maximum (FWHM) as our fiducial LOCI parameters.

As with median PSF subtraction (Section 5.1), LOCI subtracts azimuthally displaced copies of a faint source,

producing negative “wings” in the effective PSF with an integrated flux approximately equal to the flux in the core. Indeed, the integrated flux in each LOCI panel is zero to within 0.5% of the flux in the original PSFs (top-left panel). The bottom-right panel shows the effect of a random positional jitter on the effective PSFs; such a jitter could be due to unmodeled instabilities in the PSF or image registration errors. These jitters, even if only  $1/6$  of the PSF FWHM in each coordinate, smear out the PSF cores and significantly degrade the sensitivity of observations (see Figure 4), emphasizing the need for reliable, sub-pixel image registration.

Especially at small separations from the central star, LOCI suppresses companion flux more than does median PSF subtraction. This is partly because the best reference frames tend to be nearby in time (and hence

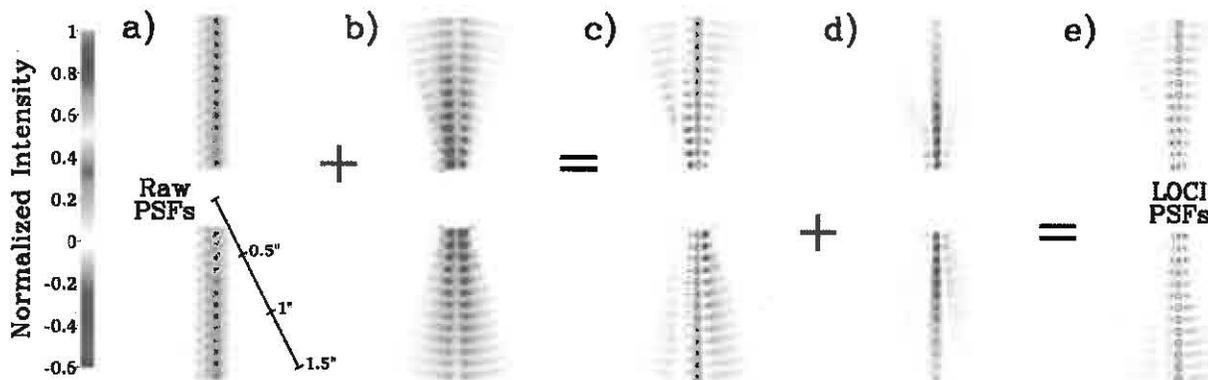


FIG. 5.— A decomposition of LOCI image processing. Panel a) shows the original PSFs, Panel b) shows the effect of subtracting angularly displaced copies of the PSF weighted by the LOCI subtraction coefficients, and Panel c) is the sum of Panels a) and b). Panel d) presents an additional source of flux suppression, which we demonstrate in Appendix A: if we perturb the flux in an image, LOCI will, on average, fit this perturbed flux with a coefficient between 0 and 1. In Panel d), this coefficient varies from  $\sim 0.5$  at  $0''.3$  to  $\sim 0.1$  at  $1''.5$ . We measure both effects as a function of spatial position and combine them to produce our sensitivity maps. Panel e), the sum of Panels c) and d), shows the effective PSFs (see Figure 4).

have little relative field rotation). Panels a)–c) of Figure 5 demonstrate this effect. Panel b) shows azimuthally-displaced PSF copies weighted by the LOCI coefficients obtained without adding a faint companion; Panel c), which closely resembles the mean-subtracted (lower left) panel of Figure 4, shows the effective PSFs after this step. Because of LOCI’s angular protection criterion, there is little intensity suppression in the PSF cores in Panel b). However, as we show in Appendix A, the addition of a faint source perturbs the LOCI coefficients themselves. To a certain degree, which varies according to AO performance and radial separation, LOCI can fit anything. Panel d) shows this effect in the same SEEDS data sequence; at small separations, it can suppress companion flux by an additional factor of  $\sim 2$ .

These two effects, described above and derived in Appendix A, account for the suppression of companion flux in a LOCI reduction. Both vary as a function of position but are nearly independent of companion flux. Figure 6 shows the results of aperture photometry on actual LOCI PSFs like those in the right panels of Figure 4. The error bars on individual points indicate the scatter in relative photometry as a function of source flux; that these are zero for the blue points (those with no positional jitter) indicates that the LOCI effective PSFs are linear in source flux. However, positional jitter, whether from AO tracking errors or from poor image registration, can introduce significant systematic errors into the recovered sensitivities.

To avoid adding test sources everywhere on the field-of-view, we produce a map of flux suppression using the method described in Appendix A. As an alternative, we could add faint sources to densely populate the FOV. However, *these sources would have to be reduced independently of one another*. If there is more than one source in an optimization region, the effect of the sources on the LOCI subtraction coefficients will change (see Equation (A6)). In general, because the linear system will be more heavily constrained, the residual intensity will be larger, and the user will overestimate his or her sensitivity.

The orange curve in Figure 6 indicates the radial pro-

file of our map of simulated flux loss, with the hatched region covering a spread of  $\pm 2\sigma$ . Because we do not compute the perturbation of the LOCI coefficients self-consistently, and because we neglect asymmetries in the companion PSF, we do not capture the full range of positional variability in relative photometry. However, our model is computationally simple and generally does an excellent job of reproducing the typical relative photometry at all angular separations; it is also free of the systematic error that we would introduce by adding and reducing many point sources simultaneously.

### 5.3. LOCI Refinements

Several authors (e.g. Marois et al. 2010; Pueyo et al. 2012; Soummer et al. 2011) have recently introduced refinements to the LOCI algorithm discussed above. These include reducing the size of subtraction regions to a single pixel, masking a small area around each subtraction region, and preconditioning the design matrix. We find that for SEEDS data, none of these steps offer a significant improvement in sensitivity. The ineffectiveness of masking an area around each subtraction region is particularly surprising. While this refinement does reduce flux suppression, it does so at the cost of additional noise. It appears that, for SEEDS data, LOCI is often as effective at fitting and removing sources as it is at fitting and removing noise. Even the subtraction of a radial profile from each image, a component of the original LOCI algorithm (Lafrenière et al. 2007b), does not improve sensitivity in typical SEEDS data.

One refinement, suggested by Marois et al. (2010) and Pueyo et al. (2012), does improve the sensitivity of some SEEDS data. Because LOCI can fit sources and noise equally well when given enough reference frames, it is preferable to reduce groups of  $\sim 100$  frames at a time; a number somewhat smaller than the size (in PSF footprints) of the optimization regions. We implement this refinement by adding the capability to process data in an integer number of groups of frames, with a single group being equivalent to a normal LOCI reduction.

We introduce three refinements of our own in addition to those listed above:

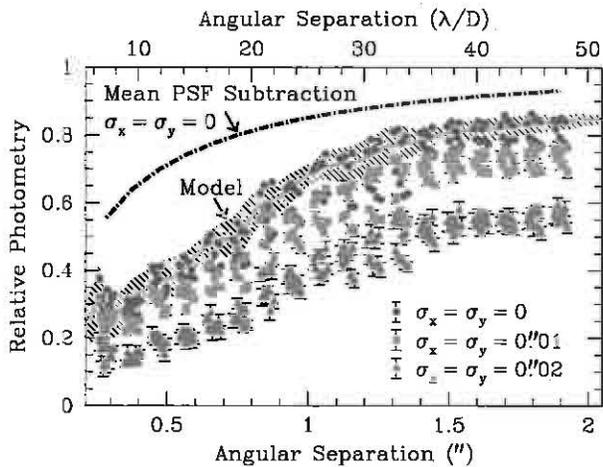


FIG. 6. — Aperture photometry of reduced point sources normalized to the photometry of the original PSF. The error bars are the standard deviations of sources of various intensities at fixed position, while the vertical scatter of points represents azimuthal variation. The effective PSF is linear in source flux (Appendix A), but varies by up to  $\sim 20\%$  with azimuthal position. The orange curve shows the radial profile of a map of fractional flux suppression computed as described in Appendix A, while the hatched region indicates  $\pm 2\sigma$  of azimuthal variation. Though it does not capture the full range of azimuthal variation, our estimated flux suppression matches the simulated sources (blue points) with no systematics. Fluctuations in the position and shape of the PSF core and image registration errors, even with standard deviations ( $\sigma_x, \sigma_y$ ) that are small compared with the PSF FWHM ( $0''.06$ ), can introduce large systematic errors in the recovered photometry.

1. Performing PCA on an ADI sequence and subtracting the first  $n$  components before applying LOCI, similar to the methods suggested by Soummer et al. (2012) and Amara & Quanz (2012);
2. Including principal components as reference frames in the LOCI process; and
3. Applying LOCI twice, to over-correct the residuals in the first application.

While we expect these refinements to be useful with a higher Strehl ratio, they seem to suppress noise and sources equally well in SEEDS data with its  $\sim 30\%$  Strehl ratio in the  $H$ -band. Soummer et al. (2012) and Amara & Quanz (2012) describe algorithms in which they use a library of PSF components like those we use for image registration (Section 4). Unfortunately, while these components are sufficiently good to register SEEDS images, they do not improve the sensitivity of LOCI. For SEEDS data, they are not even good enough to perform absolute centroiding to better than  $\sim 1$  pixel. Unlike a space telescope, HiCIAO’s AO system must be re-tuned before each observation. As a result, the PSF variation from one observation to the next is generally much larger than the variation within a single ADI sequence. Applying an initial PCA subtraction also makes it much more difficult to understand the fractional flux loss, and hence the sensitivity to point sources. In other words, it makes the flux suppression in Panel d) of Figure 5 significantly larger and harder to estimate.

We implement all of these refinements as optional features of ACORNS-ADI. With the improved performance of SCEXAO, Subaru’s next-generation adaptive optics system, or when applied to data from other instruments, these refinements may become much more powerful.

## 6. SEARCHING FOR POINT SOURCES

After reducing each frame in an ADI sequence, we combine them into a single image and search for point sources. We now discuss each step in turn. In Section 6.1, we introduce a new algorithm, intermediate between the mean and the median, to combine an image sequence. This new algorithm improves the standard deviation by up to 20% relative to taking the median of the images. In Section 6.2, we test several filters to search for point sources, settling on a  $0''.05$ -diameter circular aperture as the best choice.

### 6.1. Combining an Image Sequence

The optimal method to combine a sequence of  $N$  frames ( $N \gg 1$ ) into a final, reduced image depends on the properties of the errors in each frame. For example, if the errors are independent and normally distributed, taking the mean of all of the frames gives a combined datapoint with only  $4N/(\pi(2N-1)) \approx 64\%$  of the variance obtained by taking their median (Kenney & Keeping 1962). However, using the mean is not robust to outliers, and is a poor choice at small angular separations where speckle residuals may be highly non-Gaussian between frames.

The original LOCI algorithm (Lafrenière et al. 2007b) and various refinements (e.g. Soummer et al. 2011) simply use the median of their LOCI-processed frames, which may not be optimal for HiCIAO data, particularly in regions far from the central star where we expect read noise to dominate. We therefore use the trimmed mean, which is continuous between the mean and median: we sort the image sequence at each spatial location, and take the mean of the middle  $n$  points, discarding  $(N-n)/2$  values each at the high and low end. When  $n=1$ , this is equivalent to taking the median; when  $n=N$ , the number of images in the sequence, it is equivalent to taking the mean. We derive the efficiency of this estimator for data drawn from a normal distribution in Appendix B.

The top panel of Figure 7 shows this estimator as applied to a sample HiCIAO image sequence of 155 frames. At small angular separations, outliers are relatively common and the mean provides a poor estimator, with a standard deviation  $\sim 20\%$  higher than that of the median. Far from the central star, however, the picture is reversed; using the mean gives a large improvement in sensitivity. At nearly all separations, the optimal solution is somewhere in between, close to the median at small separations and close to the mean further away. We expect (and Figure 7 confirms) that the relationship between the optimal  $n$  in the trimmed mean and angular separation from the central star is essentially monotonic.

We implement a trimmed mean estimator iteratively. We begin with the median of our image sequence and calculate its noise profile. We then calculate the noise profile for an image created by averaging more frames (or trimming fewer), and replace data points in the median image at annuli where the new estimator reduces the

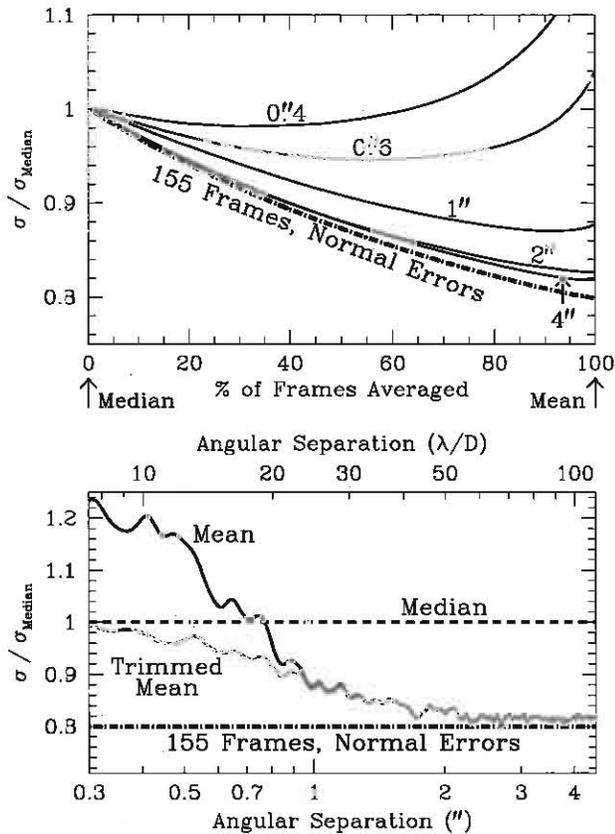


FIG. 7.— Top panel: relative efficiency of the trimmed mean (Section 6.1) as a function of the fraction of data averaged. At small separations, like  $0''.4$  (top blue curve), the median out-performs the mean, while the reverse is true far from the central star (red and bottom blue curves). However, a trimmed mean, intermediate between the mean and median, out-performs both. Bottom panel: the noise profile in a final, reduced image relative to a median combination of all frames (green dashed line). The blue curve indicated a mean of all of the frames, while the black dot-dashed line indicates the theoretical value ( $\approx \sqrt{2/\pi}$ ) for Gaussian data. The trimmed mean (red curve) out-performs the median at all radii, and by nearly 20% far from the central star.

variance. We repeat this step, using more frames (larger  $n$ ) in each successive estimator, until we use nearly all the frames. We always trim at least 5% of the data to guard against cosmic rays and rare outliers; Figure 7 shows that this approximation incurs at most a 0.5% penalty in noise.

The bottom panel of Figure 7 shows the results of our iterative trimmed mean relative to a simple median of the frames in an image sequence. The new image represents an improvement in noise at all angular separations, with a  $\sim 20\%$  improvement at large separations where the frame-to-frame noise is very nearly Gaussian.

While the frame-to-frame noise at a given pixel may be significantly non-Gaussian, the distribution of trimmed means, due to the central limit theorem, *is* Gaussian. Our new final image thus retains the noise properties of the original LOCI algorithm; in our sample dataset, it produced zero single-pixel false positives (pixels with  $>5\sigma$  fluctuations).

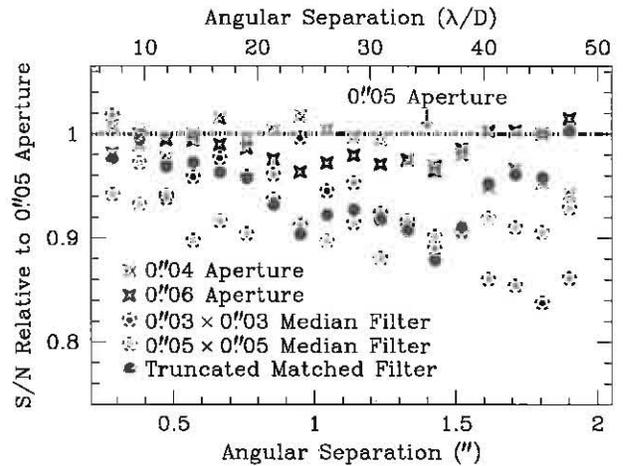


FIG. 8.— Comparison of three detection/photometry algorithms for a companion at  $1''$ ; the raw S/N is the peak companion intensity over the noise at the companion’s separation. The relative S/N is normalized to the raw S/N. PSF photometry gives slightly better results than aperture photometry, though it is best to exclude the PSF wings given the spectrum of noise. All three methods are linear to  $\sim 1\%$  over a wide range of companion fluxes.

## 6.2. Filtering the Image

The optimal filter for detecting an object depends on both the object and the character of the noise. For noise that is Gaussian and independent at adjacent pixels, the optimal filter to search for point sources is the normalized PSF, referred to as a matched filter. In this section we measure the performance of three filters for point sources in the HiCIAO data:

1. A matched filter;
2. A circular aperture;
3. A “truncated” matched filter set to zero outside an aperture; and
4. A 2-dimensional median filter.

Figure 8 shows the relative performance of these filters. We truncate the matched filter at twice the radius of our fiducial  $0''.05$ -diameter aperture to limit the impact of the outer wings, which can depend strongly on azimuthal position (see Figure 4). We do not show the full matched filter, which fails to out-perform aperture photometry even assuming perfect knowledge of the effective PSF.

Perhaps surprisingly, a  $0''.05$ -diameter ( $1.2\lambda/D$  at  $1.6\ \mu\text{m}$ ) circular aperture seems to offer the best performance of all the filters. A two-dimensional median filter is not optimal, especially far from the central star, because the noise in the reduced frame is approximately Gaussian. The poor performance of the matched filter, on the other hand, results from the strong correlation of the residual intensity in neighboring pixels. Figure 9 demonstrates this correlation in Fourier space; it results from averaging adjacent pixels when interpolating onto a new spatial array. We interpolate each frame three times during the data reduction process: when applying the distortion correction, when recentering, and finally when derotating each image to a common orientation on-sky. We have

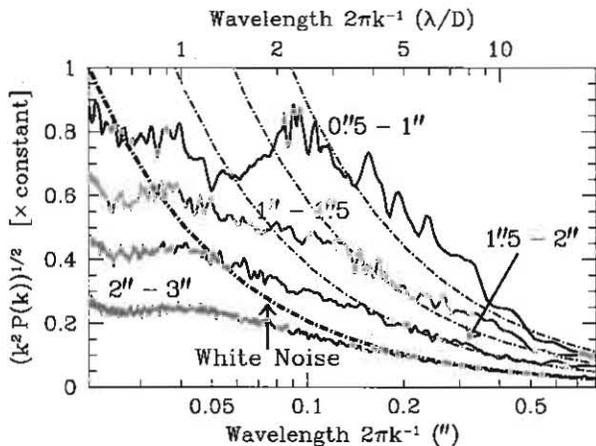


FIG. 9.— Power spectrum of noise at four annuli in our final reduced image, with white noise overplotted in black/grey. The suppression of noise at small spatial scales results from averaging. We interpolate each frame onto a new spatial array three times: during the distortion correction, for recentering, and finally when we derotate each image to a common orientation on-sky. The increase in power at a few  $\lambda/D$  is an artifact of LOCI. Because LOCI gives zero average intensity even on modest spatial scales, intensity near the central star is anti-correlated over spatial scales larger than the PSF.

verified that we can closely reproduce the power spectrum of noise at large separations by smoothing white noise.

In addition to the suppression of noise at high spatial frequency, Figure 9 shows an increase in power at a few  $\lambda/D$  (a few PSF diameters), particularly in regions close to the central star. This is an artifact of the LOCI algorithm, which tends to give zero flux averaged over a spatial region larger than a PSF core. As a result, LOCI introduces an anti-correlation in intensity over scales larger than the size of the PSF. This also helps explain the poor performance of the full matched filter, which would otherwise take advantage of the negative wings in the effective PSF: large random fluctuations will also tend to be surrounded by regions of negative intensity.

### 6.3. Producing a Sensitivity Map

To date, the vast majority of high-contrast direct imaging observations have not detected substellar companions. However, non-detections may still be used to test models of planet and brown dwarf frequency, separation, and luminosity (e.g., Lafrenière et al. 2007b, Bonavita et al. 2012). Such analyses rely on the accuracy of sensitivity maps. As we show in Figure 6, it is easy to systematically overestimate sensitivity by failing to include effects such as PSF fluctuations and image registration errors.

It has become widely accepted within the community to estimate sensitivity by computing the standard deviation in the final, combined image in annuli around the central star (e.g., Lafrenière et al. 2007b; McElwain et al. 2008; Metchev & Hillenbrand 2009; Vigan et al. 2012). To account for LOCI’s suppression of companion flux, the dataset is re-reduced after adding faint sources to compute the fractional flux loss as a function of radial separation. We begin in the same way, computing the standard deviation of the residual in-

tensity after convolving with our chosen  $0.05''$  aperture. We correct for companion flux suppression using our own method, which we describe in Section 5.2 and Appendix A.

The result of our sensitivity analysis is not simply a radial profile, but a full 2D sensitivity map, obtained with modest additional computational cost. As a final step, we multiply this map by the scaled aperture photometry of the central star in a sequence of reference images, producing a contrast map.

## 7. USE AND PERFORMANCE

ACORNS-ADI is easy to use, and requires user interaction at only two points:

1. To start the program, select the data and calibration files, and the reduction parameters; and
2. To interactively set and verify the absolute image centroid.

The total human time to perform a reduction is thus a couple of minutes, and is independent of the dataset.

The amount of computer time required scales with the size of the dataset and the number of processors available. ACORNS-ADI is efficiently parallelized, running more than 12 times as fast on 16 processors as on a single processor. A full reduction from raw data on an ADI sequence of 155 frames takes about 40 minutes on our three-year-old 16-core machine; this compares with  $\sim 2$  hours of computer time and 8 user interactions to process  $1/3$  as many optimization regions with serial IDL software adapted from Lafrenière et al. (2007b). The LOCI step scales differently depending on the number of frames. For datasets of  $\gtrsim 100$  frames, it scales with  $N_{\text{frames}}^4$ , while for much smaller datasets it scales with  $N_{\text{frames}}^2$ . The other steps in the data reduction process, with the exception of combining the images (which is computationally cheap), each scale linearly with  $N_{\text{frames}}$ .

## 8. CONCLUSIONS

We have described ACORNS-ADI, a new, parallel, open-source software package for reducing ADI data from the SEEDS survey. Most of its modules apply equally well to non-SEEDS ADI data, and the entire package could easily be adapted to analyze data from other instruments. We have introduced three new algorithms:

1. A new method of performing image registration, which is accurate to  $\sim 0.002''$ ,  $\sim 0.2$  HiCIAO pixels;
2. A new method for combining images in an ADI sequence, which reduces noise by up to 20%; and
3. A new method for calculating the flux loss in the LOCI algorithm without adding and reducing artificial point sources.

These new algorithms may be applied to any ADI dataset, improving performance and decreasing run time.

We have described and characterized each step of the ADI data reduction process for SEEDS data. With ACORNS-ADI, we will be able to process data much more quickly and efficiently, taking advantage of the SEEDS survey’s design as a large strategic observing program. In the future, we will modify ACORNS-ADI to

process data from other surveys and instruments, providing a large set of uniformly reduced data with which to perform statistical analyses of substellar companion frequencies and luminosities.

This research is based on data collected at the Subaru Telescope, which is operated by the National Astronomical Observatories of Japan. This material is

based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0646086. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

## APPENDIX

### A: PARTIAL SUBTRACTION IN LOCI

In LOCI, we build a model PSF for each frame  $I_i$  in an ADI sequence from the other frames  $\{I_j\}$  satisfying LOCI's angular displacement criterion. Denoting the intensity at pixel  $k$  in frame  $j$  by  $I_{jk}$ , we calculate the coefficients  $\{\alpha_j\}$  that minimize

$$R_i^2 = \sum_{\text{pixels } k} \left( I_{ik} - \sum_{\text{frames } j} \alpha_j I_{jk} \right)^2, \quad (\text{A1})$$

where the first sum is over pixels  $k$  in the optimization region. The coefficients  $\{\alpha_j\}$  are the solution to the linear system

$$\mathbf{A} \cdot \boldsymbol{\alpha} = \mathbf{b}, \quad (\text{A2})$$

with

$$A_{jl} = \sum_{\text{pixels } k} I_{jk} I_{lk} \quad \text{and} \quad b_j = \sum_{\text{pixels } k} I_{ik} I_{jk}. \quad (\text{A3})$$

We can perturb this problem by adding a faint source of intensity  $I'$  to frame  $i$ , and azimuthally displaced copies of it to the other frames. If there are many frames, as is almost always the case in an ADI sequence, we may approximate the azimuthally displaced copies by adding a source of intensity

$$I'_{\text{eff}} = I' - \sum_j \alpha_j I'_j \quad (\text{A4})$$

to frame  $i$ . The perturbations  $\{\beta_j\}$  in the LOCI coefficients will then be given by the solution to the linear system

$$\mathbf{A} \cdot \boldsymbol{\beta} = \mathbf{b}', \quad (\text{A5})$$

with

$$b'_j = \sum_{\text{pixels } k} \left( I'_k - \sum_{\text{frames } l} \alpha_l I'_{lk} \right) I_{jk}. \quad (\text{A6})$$

Note that the companion intensity  $I'$  and the coefficients  $\beta_j$  are both linear in the companion flux. The residual intensity in pixel  $k$  of frame  $i$ ,  $\mathcal{R}_{ik}$ , is then

$$\mathcal{R}_{ik} = I_{ik} + I'_{ik} - \sum_j (\alpha_j I_{jk} + \alpha_j I'_{jk} + \beta_j I_{jk} + \beta_j I'_{jk}) \quad (\text{A7})$$

$$\approx I_{ik} - \sum_j \alpha_j I_{jk} + I'_{ik} - \sum_j (\alpha_j I'_{jk} + \beta_j I_{jk}) \quad (\text{A8})$$

Because the source is faint, we drop the quadratic term  $\beta_j I'_{jk}$ . The first two terms in Equation (A8) give the residual intensity without the additional faint source, while the latter three terms give the residual intensity in the LOCI-processed image. These are all proportional to  $I'$ ; hence, the LOCI effective PSFs are linear in the source flux.

We compute the relative photometry of a LOCI effective PSF by evaluating the latter two terms of Equation (A8), multiplying by an aperture and summing over pixels (as described in Section 6.2, we use aperture photometry for the SEEDS data). We use a source of unit flux,

$$\sum_{\text{pixels } k} I'_{ik} a_k = 1, \quad (\text{A9})$$

where  $a$  is the aperture. Thus,  $\sum_k I'_{jk}$  is the flux in an aperture displaced from the PSF center by the relative field rotation between frames  $i$  and  $j$ . We pre-compute these fluxes as a function of position, then multiply by the LOCI coefficients  $\{\alpha_j\}$ .

We estimate the last term in Equation (A8) by first approximating  $I'_{\text{eff}}$  (Equation (A4)) as a Gaussian central peak, with several wider, negative Gaussians representing the wings (see Figure 4). We then compute  $\{\beta_j\}$  using Equation

(A5). Because we have already performed LU decomposition on  $\mathbf{A}$  to solve Equation (A2), this step takes little computational effort. We then compute the total flux loss within the aperture,

$$\sum_j \alpha_j \int_a I'(\delta\phi_j) + \sum_k \alpha_k \sum_j \beta_j I_{jk} \quad (\text{A10})$$

This allows us to compute the fractional flux loss everywhere on the FOV for little additional cost relative to LOCI itself.

## B: EFFICIENCY OF THE TRIMMED MEAN ESTIMATOR

We briefly derive the efficiency of the trimmed mean estimator used in Section 6.1 when applied to data with normal errors. The efficiency is inverse of the variance of an estimator relative to the minimum possible variance of any estimator. For Gaussian data, the mean provides the minimum possible variance, which is equal to  $\sigma^2/N$ . In the trimmed mean, we take the mean of the middle  $n$  out of a total of  $N$  points; for simplicity, we will assume  $N$  and  $n$  to be odd. We will work mostly in the space of the cumulative distribution function (CDF), in which each realization of the distribution is drawn from a uniform distribution between 0 and 1.

As a first step, we write down the probability that the middle  $n$  realizations (in CDF space) are all between  $x_1$  and  $x_2$ , with one realization each at  $x_1$  and  $x_2$  (within  $dx$ ); that is, that we have  $(N-n)/2-1$  realizations  $< x_1$  and  $(N-n)/2-1$  realizations  $> x_2$ . We denote  $(N-n)/2-1$ , the number of datapoints trimmed at each end, by  $q$ . We have

$$p(x_1, x_2) dx^2 = x_1^q (1-x_2)^q (x_2-x_1)^n \times {}_N C_q \times {}_{N-q} C_q \times (n+2) \times (n+1) \quad (\text{B1})$$

$$= x_1^q (1-x_2)^q (x_2-x_1)^n \frac{N!}{(q!)^2 n!} \quad (\text{B2})$$

Now, given  $x_1$  and  $x_2$ , we wish to calculate the variance of the mean of  $n$  realizations of the truncated normal distribution. We will assume without loss of generality that the normal distribution has zero mean. We denote these limits as  $\text{CDF}^{-1}(x_1) = \sigma y_1$  and  $\text{CDF}^{-1}(x_2) = \sigma y_2$ , where  $\text{CDF}^{-1}$  is the quartile function and  $y_2 > y_1$  are both drawn from a normal distribution with unit variance. Assuming the  $n$  realizations to be independent, the expectation value of the square of their mean is

$$\sum_{y_1, y_2} p(y_1, y_2) \left[ \frac{n}{x_2 - x_1} \int_{\sigma y_1}^{\sigma y_2} \frac{t^2 dt}{n^2 \sqrt{2\pi} \sigma^2} e^{-t^2/2\sigma^2} + \frac{n(n-1)}{(x_2 - x_1)^2} \left( \int_{\sigma y_1}^{\sigma y_2} \frac{t dt}{n \sqrt{2\pi} \sigma^2} e^{-t^2/2\sigma^2} \right)^2 \right] \quad (\text{B3})$$

We integrate the first term by parts and then integrate the second term. The inverse of the efficiency is then  $\sigma^2/N$  times Equation (B3), equal to

$$\frac{N}{n} \sum_{y_1, y_2} \frac{p(y_1, y_2)}{x_2 - x_1} \left[ \frac{y_1}{\sqrt{2\pi}} e^{-y_1^2/2} - \frac{y_2}{\sqrt{2\pi}} e^{-y_2^2/2} + \int_{y_1}^{y_2} \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} + \frac{n-1}{2\pi(x_2 - x_1)} \left( e^{-y_1^2/2} - e^{-y_2^2/2} \right)^2 \right] \quad (\text{B4})$$

$$= \frac{N}{n} \int_0^1 dx_1 \int_{x_1}^1 dx_2 \frac{p(x_1, x_2)}{x_2 - x_1} \left[ \frac{y_1}{\sqrt{2\pi}} e^{-y_1^2/2} - \frac{y_2}{\sqrt{2\pi}} e^{-y_2^2/2} + x_2 - x_1 + \frac{n-1}{2\pi(x_2 - x_1)} \left( e^{-y_1^2/2} - e^{-y_2^2/2} \right)^2 \right] \quad (\text{B5})$$

We then substitute for  $p(x_1, x_2)$  using Equation (B2) and evaluate the integral. In the limit of the median ( $n=1$ ), Equation (B5) reduces to  $\pi(2n-1)/4n$ , for an asymptotic efficiency of  $2/\pi \approx 64\%$  (Kenney & Keeping 1962).

## REFERENCES

- Alonso, R., Brown, T. M., Torres, G., et al. 2004, *ApJ*, 613, L153  
 Amara, A., & Quanz, S. 2012, *ArXiv e-prints*, 1207.6637  
 Bakos, G., Noyes, R. W., Kovács, G., et al. 2004, *PASP*, 116, 266  
 Bertin, E., & Arnouts, S. 2010, *Astrophysics Source Code Library*, 10064  
 Biller, B. A., Close, L. M., Masciadri, E., et al. 2007, *ApJS*, 173, 143  
 Blank, R., Anglin, S., Beletic, J. W., et al. 2011, in *Astronomical Society of the Pacific Conference Series*, Vol. 437, *Solar Polarization 6*, ed. J. R. Kuhn, D. M. Harrington, H. Lin, S. V. Berdyugina, J. Trujillo-Bueno, S. L. Keil, & T. Rimmele, 383  
 Bonavita, M., Chauvin, G., Desidera, S., et al. 2012, *A&A*, 537, A67  
 Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, 327, 977  
 Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, *Nature*, 462, 891  
 Egner, S., Ikeda, Y., Watanabe, M., et al. 2010, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7736, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*  
 Guyon, O., Martinache, F., Clergeon, C., et al. 2011, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8149, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*  
 Hayano, Y., Takami, H., Guyon, O., et al. 2008, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7015, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*  
 Janson, M., Bonavita, M., Klahr, H., & Lafrenière, D. 2012, *ApJ*, 745, 4  
 Kenney, J. F., & Keeping, E. S. 1962, *Mathematics of Statistics*, Pt. 1, Third Edition (Princeton, NJ: Van Nostrand), 211  
 Lafrenière, D., Doyon, R., Marois, C., et al. 2007a, *ApJ*, 670, 1367  
 Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007b, *ApJ*, 660, 770  
 Liu, M. C. 2004, *Science*, 305, 1442  
 Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, *ApJ*, 641, 556

- Marois, C., Macintosh, B., & Véran, J.-P. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7736, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, *The Messenger*, 114, 20
- McCullough, P. R., Stys, J. E., Valenti, J. A., et al. 2005, *PASP*, 117, 783
- McElwain, M., Larkin, J., Metchev, S., & Zuckerman, B. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Metchev, S. A., & Hillenbrand, L. A. 2009, *ApJS*, 181, 62
- Moseley, S. H., Arendt, R. G., Fixsen, D. J., et al. 2010, *Proc. SPIE*, 7742
- Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. 2006, *PASP*, 118, 1407
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, *Numerical Recipes, Third Edition* (New York, NY: Cambridge University Press)
- Pueyo, L., Crepp, J. R., Vasisht, G., et al. 2012, *ApJS*, 199, 6
- Queloz, D., Mayor, M., Weber, L., et al. 2000, *A&A*, 354, 99
- Soummer, R., Brendan Hagan, J., Pueyo, L., et al. 2011, *ApJ*, 741, 55
- Soummer, R., Pueyo, L., & Larkin, J. 2012, *ArXiv e-prints*, 1207.4197
- Tamura, M. 2009, in American Institute of Physics Conference Series, Vol. 1158, American Institute of Physics Conference Series, ed. T. Usuda, M. Tamura, & M. Ishii, 11
- Thalmann, C., Janson, M., Buenzli, E., et al. 2011, *ApJ*, 743, L6
- Tinney, C. G., Butler, R. P., Marcy, G. W., et al. 2001, *ApJ*, 551, 507
- Vigan, A., Patience, J., Marois, C., et al. 2012, *A&A*, 544, A9
- Vogt, S. S., Marcy, G. W., Butler, R. P., & Apps, K. 2000, *ApJ*, 536, 902